PC Support Advisor

The Essential Resource for PC Support Professionals

Update 73

How To Troubleshoot Memory Errors
One lost bit can spell disaster....................3
Julian Moss on how to tell when something's not quite right with your RAM, and what to do about it.
File: H0722.1

Understanding Systems Procurement
How to get what you need, at the best price....................7
Dave Stott presents the first of a two-part article on systems procurement, showing how to undertake the process in the best way possible and how to get the best prices.
File: M0126.1

Understanding PC Buses
PCI, VLB, EISA, ISA, MCA - what does it all mean? 11
Julian Moss concludes his article on the ins and outs of PC bus architectures.
File: H0820.9

From OS/2 To Warp
Does IBM really stand a chance? 15
Just what is OS/2, and is it better than Windows for your workstations or servers? Terence Green has the answers.
File: O0417.1

Understanding Printer Technology
Laser, inkjet, dot matrix and the rest explained....................19
Chris Long presents an overview of some of the most common printing technologies found on today's desktops.
File: H0821.1

On This Month's Utility Disk

Windows utility to locate duplicate files on a hard disk.

Utility to maximise amount of memory available for Windows.

Installation application generator. Useful for shipping files and updates to users.

"Beware Geeks Bearing Gifs" - says Robert Schifreen.

It's impossible to identify just one single reason for the success of Windows 3.0. But one important factor was certainly the way that Microsoft announced the name of the product months ahead of shipping any actual code. Officially, the name was secret, but Microsoft didn't seem to mind journalists and beta testers giving away snippets of information.

In the months before the launch of 3.0 there were literally hundreds of magazines reviewing hardware and software which "worked perfectly with a certain non-existent operating system" or "ran version 3 of a well-known software package very well".

The whole thing was a wonderful marketing ploy. It made the readers feel unhappy because the magazines had copies of Windows 3.0 but they didn't. So, naturally, as soon as Windows 3.0 hit the shelves, everyone rushed to buy it.

Microsoft now seems to have abandoned this form of marketing stunt. Having watched public awareness of the names "Chicago" and "Windows 4.0" reach a massive high, Microsoft has formally announced that this product will actually be called Windows 95. Whether it will actually ship in '95 remains to be seen. What does make good sense, though, is that, in 1996, no-one will want to be seen using software which is so obviously out of date. So perhaps Microsoft has got it right yet again.

On another subject entirely, the world's newspapers and magazines (not just the computer ones) have jumped on the internet bandwagon. Unfortunately, a network that allows millions of computer users around the world to send messages and files is not exactly headline news. But a system which allows a child in one corner of the world to download pornographic images from a computer in another makes a great story. So, computer porn stories are becoming more and more popular.

In London, England, the BBC has a collection of Internet web pages which provide details of the week's TV programmes. There are also pages of information for children. Yet one British newspaper recently accused the BBC of providing a service on which kids could download pornographic images!

Where will all this end? A colleague of mine suggested that perhaps I should prosecute my local council for building roads linking my office to a downtown porn shop.

I've talked about the subject before, but it's worth repeating the warning again. Keep a close watch on the contents of your users' hard disks and the corporate LAN servers. Yes, there are some pretty nasty graphic files (still images and video clips) available on some of the seedier internet sites. Having your company's computers exposed as a source or possessor of such files is an excellent way to lose your job and most of your clients, so it pays to be vigilant.

PCSA

PC Support *Advisor*

PC Support *Advisor* is a publication of **International Technology Publishing**.

ISSN: 1031-3966

Editor: Robert Schifreen

Editor in Chief: Les Bell

Publisher: Chris Carvan

Email:
pcsa-ed@oakworth.demon.co.uk

Copyright:
All material is *Copyright © 1994 International Technology Publishing*, and must not be reproduced in part or full, by any means, without the written permission of the publisher.

Liability:
The information, opinions, interpretations and directions herein are based on the best information available, but their completeness and accuracy cannot be guaranteed.

Responsibility:
Responsibility for the views and opinions expressed in this publication lies with the publisher, International Technology Publishing, and not with any agents of the publisher nor the editors or contributing editors.

International
Technology
Publishing Inc.
North American Subscriptions
Suite 200
555 De Haro Street
San Fransisco, CA94107
USA

Telephone: (415) 255 1295
Facsimile: (415) 255 8496

International
Technology
Publishing Inc.
European Subscriptions
13 Callcott Street
London
W8 7SU
United Kingdom

Telephone: (0171) 724 9306
Facsimile: (0171) 706 3296

International
Technology
Publishing Pty Ltd.
SE Asian Subscriptions
Level 36 Hong Leong Blding
16 Raffles Quay
Singapore
0104

Telephone: (65) 321 8984
Facsimile: (65) 322 8558

International
Technology
Publishing Pty Ltd.
Australasian Subscriptions
Suite 511
185 Elizabeth Street
Sydney, NSW, 2000
Australia

Telephone: (02) 261 4683
Facsimile:(02) 261 4741

How To Troubleshoot Memory Errors

Julian Moss on how to tell when something's not quite right with your RAM, and what to do about it.

Reliable memory is critical to the operation of a PC. Faulty memory can result in corrupted data, or may cause the system to crash. It only needs a single bit to change from its correct value to change the meaning of a field in a database record or a cell in a spreadsheet, or to alter a program instruction to something the programmer never intended.

There are over 64 million bits stored in the memory of a typical PC with 8 MB of RAM, and to expect every one of them always to be correct is to expect a high degree of reliability. Unfortunately, that expectation is not always met.

Missing RAM

One of the common causes of memory errors is a poor electrical contact. This usually results in the symptom of "missing memory", when the system reports that it has less RAM than you know is physically installed in the machine. This is usually noticeable at boot-up, when the BIOS power-on self-test (POST) memory checks are carried out.

The bad contact makes it seem as if no memory exists at a particular location. When the POST checks reach a memory address that doesn't appear to exist, the BIOS assumes that it has come to the end of available memory, and uses the figure it has counted up to as the amount of memory in the system.

This fault is very easy to cure. In the case of SIMMs, all you normally need do is to remove each module and then replace it. Friction between the contacts is usually enough to clean them and ensure that a good connection is made. Often, the problem is found to be caused by the module not being seated properly in its socket in the first place.

If a PC is kept in a humid or dirty environment, it's possible that the contacts on the SIMMs and their sockets have become contaminated. Any dirt, dust or grease can be removed by applying electrical contact cleaner - available from component suppliers and Tandy/Radio Shack stores - in accordance with the maker's instructions.

Better quality SIMMs have gold-plated contacts, so oxidation should not be a problem. Never use anything abrasive to clean the contacts, as the gold plating is very thin and will be scraped off.

In older PCs that use dual-inline-package (DIP) memory chips, bad contacts can be caused by the chips creeping out of their sockets as a result of thermal expansion and contraction over a period of time. This applies to socketed SRAM used in the memory cache of current systems, too.

The solution here is to press each chip firmly back until it is correctly seated in its socket. You'll hear some fearsome cracking sounds when you do: this is normal.

A conscientious field service engineer will remove the motherboard and place it on a flat surface before doing this, to avoid flexing the board too much. However if the motherboard is not too cheap and flimsy and has been properly mounted you should be able to press the chips in far enough to cure the problem with the board *in situ*. Just don't press too hard.

Parity Errors

Other types of memory fault can be more difficult to detect. A problem may be intermittent, or it may appear to be intermittent because it only shows up when a particular memory location is accessed and certain values are stored in it.

A single bit error is the most common type of memory error. This can be detected by the parity checking hardware built into every motherboard,

> *"If the electrical supply is of poor quality, high voltage spikes may travel through the PC's power supply and appear on the +5VDC rail. These spikes can damage components. If PCs are suffering frequently from memory faults, have the supply checked out."*

but only if memory with 9 bits per byte (8 data bits plus a parity bit) is used. When a byte is written to RAM, the parity circuits set the value of the parity bit so that the total number of ones is even. When a memory location is read, the parity is checked. If it is incorrect - in other words, the number of ones is odd - bit 7 is set in I/O port 61h and a non-maskable interrupt (NMI) is generated.

The NMI is serviced by a routine in the BIOS. This checks the status of I/O port 61h to see what type of error has caused the NMI: other fatal errors are also reported in this way. If bit 7 is set, the BIOS attempts to generate the parity error again so that it can display the address of the location where it occurred.

If the error occurred in the system board memory, then you will see a message similar to:

```
Memory Parity Error at xxxxx
```

where xxxxx is the address in hex of the faulty location. This is the message reported by the Hi-Flex AMI BIOS; others may use a slightly different message. If the error is in memory on an expansion card, you will see the message:

```
I/O Card Parity Error at
xxxxx
```

If the BIOS is unable to find the error, then question marks are displayed in place of a memory address: this could signify an intermittent error or other motherboard fault.

Note that parity can only detect a single bit error, and it cannot correct it. High end PCs designed for server applications may use error correction code (ECC) memory. Instead of parity bits, ECC uses six or seven dedicated error correction bits per 32-bit word. This enables single bit errors to be corrected, so the system doesn't have to be brought to a halt. Errors of two bits or more - which a parity check might not spot - will be detected.

Undetected Errors

Windows and other protected mode software is more likely to show up memory faults because it makes use of all the memory. DOS only makes really intensive use of the bottom 640 KB.

A memory error, if undetected, could lead to corrupt data being written to disk (from a disk cache buffer, for example) as well as unexplained system crashes. So it is worrying that many PCs are now supplied with memory that does not have a parity bit, and with the parity detection circuits disabled. This trend is more common at the cost-sensitive bottom end of the market, but is by no means confined to it.

The memory test performed during the POST is nowhere near thorough enough to show up all types of memory fault. If it was, it would take an intolerably long time. In many BIOSes it can in any case be turned off altogether. So many PCs run without ever testing whether the integrity of the RAM is good or not.

Microsoft believes that undetected memory errors may be responsible for some of the problems users have experienced with DoubleSpace as well as unexplained Windows crashes. Cynics may note that Microsoft never admits to bugs in its software. Nevertheless, it has attempted to prevent the problem by having HIMEM.SYS perform a memory test when it loads. This was introduced with MS-DOS 6.20. However, HIMEM's checks aren't infallible either because it simply isn't possible to test memory thoroughly in a couple of seconds.

Testing Cache And Shadow RAM

Cache RAM and ordinary RAM used to shadow BIOS ROMs are difficult to test. Often, a fault here will result in a system that won't boot up at all. But if the fault is intermittent, or lies at an address that is infrequently accessed, it may be responsible for random system crashes and other unexplained faults.

Shadow RAM is used to speed up access to the data and program instructions held in the BIOS ROM. 386 memory managers can be used to shadow a ROM, but usually this is done in hardware by the PC's chip set. While ROM is being shadowed, the memory being used to shadow it is inaccessible to any memory diagnostic unless it knows enough about the chip set to overcome this. DiagSoft's QAPlus/FE can test shadow memory, but only on a very limited number of systems.

Although your diagnostic software may not have the capability to test the shadow RAM in your system, it can test the memory if you give it some help. Go into the BIOS Setup utility, and look for either the Advanced Setup or Advanced Chipset Setup screens. If neither is present, shadowing is probably enabled automatically and you'll need the manufacturer's diagnostics disk to test the memory used.

If your BIOS does give access to the Advanced Setup screens, look for an option to enable shadowing. In some BIOSes there is a list of options for shadowing blocks of ROM from C0000h to the top of base memory. Make a note of the current settings and then disable all shadowing. The shadow RAM will now be made available as ordinary user-accessible memory, and can be tested by the diagnostic software. System I/O performance will be affected, as the BIOS must now be read from slow ROM. But this won't affect the time taken by the memory test.

Cache memory is a more difficult problem. It can't be tested, except by manufacturer-produced diagnostics. However, if a system has a fault in its secondary cache, it isn't likely to get very far before failing. If you suspect faulty cache RAM is at the root of a problem, try disabling it from the Advanced Chipset Setup screen. The system will run slower, but if it runs without error then you can be fairly certain that you've found where the problem lies.

Memory Errors

Types Of Memory Fault

The main memory in a PC is Dynamic Random Access Memory (DRAM). Each bit value is stored not in a semiconductor switch but in a capacitor. A binary 1 is indicated by the presence of a charge, which starts to decay from the moment it is received. Consequently, DRAM has to be "refreshed" periodically, to restore the charge before it decays below a certain level.

The refresh period is about 15 microseconds. This is more than frequent enough to retain the information in correctly functioning RAM. However a weak memory cell may not hold its data, and so a parity error will occur.

Other categories of memory failure that can occur are shown in Figure 1. The nature of some of these types of failure are such that an error may not occur every time a memory location is accessed. The fault will appear to be intermittent. Such faults can be difficult to locate, even using diagnostic software.

> *"When the POST checks reach a memory address that doesn't appear to exist, the BIOS assumes that it has got to the end of available memory, and uses the figure it has counted up to as the amount of memory in the system."*

Static Electricity

Static electricity can cause memory faults to develop, by damaging or weakening the semiconductor circuits within the memory chips. You can avoid the possibility of damage by taking anti-static precautions when handling memory modules. Work on an anti-static mat, and use a wrist-strap to ground yourself to the PC chassis. Hold SIMMs by the edges and DIPs by the epoxy package; don't touch the electrical contacts.

If the electrical supply is of poor quality, high voltage spikes may travel through the PC's power supply and appear on the +5VDC rail. These spikes can damage components. If PCs are suffering frequently from memory faults, have the supply checked out by a qualified power consultant, and if necessary install power conditioners.

Don't mix memory modules of different speed ratings - or even those with identical ratings but made by different manufacturers - within a memory bank. Slight differences in the timing signals from the different chips may upset the memory controller circuits, resulting in apparent memory errors.

Testing Memory

If you suspect that a PC has a memory fault, you can verify this using diagnostic software. Suitable programs include QAPlus and QAPlus/FE by DiagSoft, TouchStone Software's CheckIt PRO, NDIAGS from the Norton Utilities and RAMexam by Qualitas. Most of these programs use tests such as Checkerboard, where memory is filled with 01010101 and 10101010 bit patterns, and walking bit tests where memory is filled with 00000001, then 00000010 and so on. These tests will find stuck bits and certain other types of fault, but their efficacy cannot be mathematically proven.

Qualitas claims that its RAMexam is more effective because it uses a consistent underlying fault model designed to detect the categories of error listed earlier. For more background, refer to "Testing Semiconductor Memories: Theory and Practice" (A J van de Goor, John Wiley & Sons, 1991).

When To Test

When to run memory tests is another point to consider. RAMexam is designed to be run on a regular basis and can perform various levels of test at daily, weekly, monthly or user-defined intervals. The $29.95 program is available on site licence terms. If your PCs have memory that is not parity-checked then this is something you may want to consider. However, even a moderate level of checking will take a minute or two, which users may find unacceptable.

The other point to bear in mind is that running a memory test on a normally-configured system is not a particularly good idea. It is not usually

Stuck-At Fault - A bit retains a value, regardless of what is written to it.

Transition Fault - Once changed, the value of a bit becomes stuck-at.

Unlinked Inversion - A bit changes erroneously as a result of a change to a bit elsewhere in memory.

Linked Inversion - A bit changes erroneously as a result of changes to two or more bits elsewhere in memory.

Unlinked Idempotent - A bit becomes stuck-at as a result of a change to a bit elsewhere in memory.

Linked Idempotent - A bit becomes stuck-at as a result of changes to two or more bits elsewhere in memory.

Figure 1 - Categories of memory failure

> *"In some BIOSes there is a list of options for shadowing blocks of ROM from C0000h to the top of base memory. Make a note of the current settings and then disable all shadowing. The shadow RAM will now be made available as ordinary user-accessible memory."*

possible to test memory used by TSRs, device drivers and so on (programs that try often lock up the system) so you aren't performing a very thorough test.

If a 386 memory manager is loaded then the memory addresses seen by the diagnostic software may not be the same as the physical memory addresses, due to the processor's memory paging mechanism. Consequently the output from the diagnostics will probably be inaccurate.

What you should do is boot the system from a floppy disk that contains an empty CONFIG.SYS and an AUTOEXEC.BAT that merely runs the diagnostic utility. Norton's NDIAGS has a comprehensive memory test that automatically reboots the PC with a minimum configuration, which makes things easy. Note that the comprehensive test takes typically well over an hour to run.

Identifying Faulty Chips

Older PCs had memory made up of columns of nine DIP memory chips (eight data bits and one parity). Each chip stored one specific bit for a range of addresses, so identifying the faulty one was quite tricky. Software such as QAPlus will do the job for you, if you tell it the layout of the memory chips on the motherboard. You still need to know which rows of chips hold bits 0 to 7 and which row is the parity row, as well as how the byte-wide columns are organised into banks. This information may be printed on the motherboard, or given in the manual.

If you don't have this information, identifying which chip is at fault is quite difficult. Trying to pinpoint the problem by swapping chips when there may be 36 of them on the motherboard is not to be recommended. Fortunately most of the systems you will be supporting today use SIMMs.

Old-style 30-pin SIMMs are one byte wide, so you don't need to worry about which bit is faulty. All you need to do is divide the address of the error by the number of SIMMs in a bank: the remainder will tell you which SIMM within a bank must be changed. Of course, you still need to know how the SIMMs are organised within a bank.

If more than one bank of SIMMs is filled, you also need to know how the banks are used. If two banks are filled, some motherboards interleave them - effectively using them as one bank of double the width - to improve memory access times. Where this is the case (check the system manual and the BIOS Setup) double the number you divide by.

Another obstacle to identifying the faulty memory module is that the addresses of the memory within a bank may not be what you think. For example, a bank containing 4 MB of RAM may actually have addresses from 0 to 4480 KB. The reason is that the chipset changes the addresses of memory above 640 KB so that no system RAM is wasted overlapping the 384 KB ROM/adapter RAM area.

The starting address of bank 1 will be (Size of memory bank 0)KB + 384 KB - (Amount of RAM used for ROM shadowing)KB. You will probably need to refer to the BIOS Advanced Setup screen to find out the last figure. However, you only need bother about it if the address of the fault is close to the starting address of a memory bank.

PS/2 Modules

The same logic can be used to identify faulty SIMMs in banks of 72-pin PS/2 type modules. Remember here that the modules are 4 bytes wide. In a Pentium system you divide the address by 8, because you have a 64-bit data bus. If the remainder is 0 - 3 then the fault is in the first SIMM of the pair, if it is 4 - 7 it is in the second.

It sounds complicated, but the hardest part is usually obtaining the information about how the SIMM banks are organised. If you have the data and a good memory diagnostic package, troubleshooting memory errors is easy.

PCSA

The Author

Julian Moss is a PC consultant and writer. He can be contacted via pcsa@oakworth.demon.co.uk

Understanding Systems Procurement

Dave Stott presents the first of a two-part article on systems procurement, showing how to undertake the process in the best way possible and how to get the best prices.

The process of purchasing a large computer system or even a number of PCs can often be quite traumatic. Not only do you have to ensure that the system(s) purchased properly fit your requirements, but in the interests of economy you have to try to make sure you are getting true value for money. If you are working to a budget then the financial aspect can be a vital component in the success of a project.

In this two-part feature I will help you to approach the procurement of computer systems in a methodical and consistent manner. I will start off by detailing a formal method of evaluating proposals which should make the task of determining the suitability of a particular system much easier. Once the desired system(s) has been identified I will proceed, in part two, with the task of negotiating the best possible deal with our chosen supplier.

Let us assume that you have prepared a detailed specification of your requirements and that you have invited several suppliers to tender for the business. As a result you will no doubt receive a number of different proposals offering a variety of possible solutions. So how do you establish which proposals are valid and, more importantly, which best suit your particular needs both technically and financially.

Approach

The approach I would recommend is based on a formal method of evaluation known as Weighted Ranking By Levels (WRBL). Often used by computer consultants to formally evaluate suppliers' quotations or proposals, this method can help you sort the wood from the trees and make objective decisions as to whether a particular system meets your needs better than another. Equally important, using this approach can help you decide which system really offers the best value for money.

In order to achieve consistency it is important to try to be objective in your decision making process. Therefore you must disregard subjective values such as the practically meaningless advertising statements like; "Small Footprint", "Lightweight", "Top of the Range" or "Our Fastest Ever System" in favour of more measurable factors such as Processor Clock Speed, Hard Disk capacity and RAM expansion capability. The simple rule is if you can't measure or quantify a specific element then just forget about it with regard to the formal evaluation process.

Initial Steps

Before starting on the task of evaluating a proposal you must first be aware of your specific requirements. These are generally determined by whatever the system(s) will be used for and presumably you will have defined these at the outset before inviting proposals from suppliers. First of all you need to decide which features of the system(s) affect how well those specific requirements are met. Basically you need to define a yardstick against which competing systems can be measured in accordance with your needs. For example; Do you foresee the need for future expansion? Is bundled software important to you? Will the system be fast enough? Will it fit into the available space? Is the system light enough to be easily portable? These are just some of the questions which may or may not be pertinent to your particular requirements.

Draw up a list of your requirements in order of priority and, from this, try to establish the critical components or features which directly relate to them. For example, if your requirement is for a system to run Computer Aided Design (CAD) then a large high-resolution monitor and a fast processor are likely to be significantly important on your list. However, if you just need a system for standard word processing then a more modest system configuration would probably suffice. The actual priority of particular aspects of the system are your own concern and they may differ significantly in various scenarios.

Whatever your requirements are, try to ensure that you cover every aspect of the system that is important to you. If, for instance, the physical dimensions need to be considered then don't forget to include these as a requirement. Similarly the non-physical aspects of a system such as software, warranty, training and technical support may be very important factors. Some aspects may be difficult or even impossible to quantify effectively in which case you may be forced to ignore them. However, it is usually possible to set some form of objective criteria by which a particular element may be judged. For example, the cost of a days training or the length of time for which free technical support is offered could be used as an objective measurement.

Benchmarks are often used to evaluate systems but they can pose a bit of a problem at times. To start with some suppliers are extremely unwilling to provide benchmark information about their systems whilst others are more co-operative. However, the main difficulty is making sure that the same benchmark (even the same version of benchmark) has been used by each supplier in order to establish a fair comparison. Naturally if you can perform your own set of benchmark tests in-house then you can be more sure of the validity of the tests and their fairness, and therefore you can safely include the results in your evaluation.

Once your list of the most important features is complete you should group them under related areas. For example, the system unit section should identify features such as the processor speed, hard disk factors and expansion capabilities. The monitor section will detail things like the resolution, the refresh rate and the physical display area.

Importance

The next task is to decide on the relative importance of each main group and each detailed feature within a particular group. For instance, if you have four main groups then these would be classed as the first level on our WRBL model and their importance relative to one another will determine the associated weighting factor. Quite simply, if you consider the screen to be equally important as the system unit then they will each have the same weight. But if you feel that the screen is only half as important as the system unit then it should be given half the weight of the former and so on - see Figure 1.

Each main group in the model may be further subdivided into a secondary level group which is used to define finer details for evaluation. In our example I have split the level one SYSTEM UNIT into further groups at level

								SUPPLIERS			
LEVEL	REQUIREMENTS	WEIGHTS				RAW VALUES			WEIGHTED VALUES		
		1	2	3	WT	A	B	C	A	B	C
1	SYSTEM UNIT	60									
2	BASE CONFIGURATION		70								
3	CPU Clock Speed in MHz/10			20	8.4	6.6	6	6.6	55.44	50.4	55.44
3	RAM in MB/10			25	10.5	1.6	0.8	1.6	16.8	8.4	16.8
3	Cache SRAM in KB/100			15	6.3	2.56	5.12	5.12	16.128	32.256	32.256
3	Video Memory in MB			2	0.84	1	1	2	0.84	0.84	1.68
3	Hard Disk capacity in MB/100			10	4.2	5.2	4.5	4.1	21.84	18.9	17.22
3	Hard Disk average access time (ms)			15	6.3	-10	-12	-11	-63	-75.6	-69.3
3	Serial port type			8	3.36	1	2	1	3.36	6.72	3.36
3	Parallel port type			5	2.1	2	2	1	4.2	4.2	2.1
				100					55.608	46.116	59.556
2	EXPANSION		30								
3	ISA 16-bit slots (free)			10	1.8	4	4	5	7.2	7.2	9
3	PCI slots (free)			15	2.7	3	1	2	8.1	2.7	5.4
3	3.5in drive bays (free)			15	2.7	2	2	1	5.4	5.4	2.7
3	5.25in drive bays (free)			35	6.3	4	3	2	25.2	18.9	12.6
3	Maximum RAM capacity			25	4.5	6.4	12.8	6.4	28.8	57.6	28.8
			100	100					74.7	91.8	58.5
1	MONITOR	20									
2	Dot pitch (value -1)		60		12	0.74	0.74	0.72	8.88	8.88	8.64
2	Maximum refresh rate in MHz/10		40		8	10	9	9	80	72	72
			100						88.88	80.88	80.64
1	SOFTWARE	15									
2	Windows NT		50		7.5	1	1	0	7.5	7.5	0
2	DOS		15		2.25	1	1	1	2.25	2.25	2.25
2	Windows 3.11		15		2.25	1	1	0	2.25	2.25	0
2	Applications bundle		20		3	4	3	3	12	9	9
			100					24	21	11.25	
1	WARRANTY	5									
2	Full refund (days/10)		25		1.25	3	1.5	2.1	3.75	1.875	2.625
2	On-site (months/10)		50		2.5	1.2	1.2	2.4	3	3	6
2	Return to base (years)		25		1.25	1	1	3	1.25	1.25	3.75
		100	100						8	6.125	12.375
								Totals	251.188	245.921	222.321

Figure 1 - Weighted Ranking by Levels example

Systems Procurement

> *"Before starting on the task of evaluating a proposal you must first be aware of your specific requirements. These are generally determined by whatever the system(s) will be used for."*

two covering the BASE CONFIGURATION and EXPANSION. These have then been further divided to show the actual details at level three. So under EXPANSION, for example, I have used the number of free 16-bit ISA slots as a specific measurement of our requirements. In theory you can have as many levels as you wish. However, you should restrict yourself to three levels as this will avoid over complicating things and will provide sufficient detail to cover most normal situations.

Employing A Spreadsheet

As you can see from our example this method of evaluation is an ideal candidate for employing a spreadsheet. It doesn't matter which actual spreadsheet package you use for this exercise as the calculations involved are fairly simple and any decent spreadsheet package will do the job. But the use of a spreadsheet is strongly advised as it enables you to apply "What If?" techniques to the evaluation exercise. If you need to adjust the weighting of a particular element at any time then this can easily be done if you are using a spreadsheet as opposed to a paper model. For example, you may find that for a specific situation "Bundled Software" is very important and giving it a higher weighting will place more emphasis on this aspect of the system with the end result being that a particular system comes out as more favourable than previously envisaged.

Once all the levels and the individual items have been defined you can determine and assign the associated weights that will apply to each level and each item within a level. Note that the weights for any particular level should add up to 100 and in our example I have used control totals to ensure that the spreadsheet is correct in this respect. The column labelled WT is a figure calculated from the weight given to each level and is used to influence the RAW VALUES to produce the WEIGHTED VALUES for the elements of each suppliers system being evaluated.

The actual calculation for the WT column is merely the weight at every superior level multiplied together and then divided by an appropriate factor to produce a suitable figure to act as an acceptable influence. For example in our spreadsheet the weight that affects the CLOCK SPEED is produced by considering three levels with the formula (60 x 70 x 20)/10000 giving a figure of 8.4, whereas the weight for MAXIMUM REFRESH RATE is calculated using only two superior levels with a formula of (20 x 40)/100 giving a weight of 8.

This approach ensures that no single weight is too significant and liable to distort the final totals. The design of the method means that the weight has an equal affect on each result.

Entering The Values

Although much of the information required for this exercise can be gleaned from advertisements or recent product reviews in magazines it is better to extract the data from actual proposals or quotes from prospective suppliers. Occasionally you may need to refer to manufacturers' technical specification data sheets or talk directly to a supplier to determine specific facts.

The most important thing to remember when entering the individual values is that they should all affect the results by influencing them in the same direction, either upwards or downwards. Most models are designed on the basis that the higher the individual value the better the overall result. This may require some thinking about with certain elements where the obvious value suggested may have the opposite effect and our example shows two different ways of rectifying this situation.

For an element such as "hard disk average access time (ms)" it is necessary to use negative values, as the lower the average access time of a hard disk is in milliseconds, the better the performance will be. If you use the straightforward positive figures the effect will be the exact opposite of what is required. This is in contrast to most other values you will enter whereby the higher the value the better the result. Therefore, by using a negative RAW VALUE you will achieve the desired effect on the WEIGHTED VALUE of this particular element.

Similarly with the "dot pitch (value -1)" element, the smaller the dot pitch on a monitor the better the overall resolution achievable and therefore you must adjust the actual value supplied by the manufacturer by subtracting it from 1. So that, for instance, a dot pitch of 0.31mm becomes a value of 0.69 and all the values are influenced in the right direction.

Where the evaluation of a particular element can only be regarded in terms of yes or no, as with the inclusion of specific SOFTWARE in our example, then you only need to assign a value of zero or one for it to have the desired effect on the result. A negative

Supplier	System Cost	Points per $100
A	$1,495	1.68
B	$1,295	1.90
C	$1,295	1.72

Figure 2 - Who offers the best deal?

> *"The method can be used to evaluate any number of requirements from any number of suppliers and the more complicated the requirement, for example for a Local Area Network, the more useful this approach."*

value will have a detrimental effect on a suppliers result and a zero value will merely have no effect. In our example the "applications bundle" has been scored on the basis of the number of packages included in the deal.

The magnitude of the figures used for the RAW VALUES is not strictly important as the totals should be viewed in terms of the ratios between suppliers overall results. However, it's best to apply some form of scaling to the figures used so that you do not end up with unwieldy numbers. To this end you will see that I have used different units of measure and scaling within the WARRANTY group to achieve the effect required.

There is nothing wrong with using either negative figures or modified scaling in the calculations so long as your approach is consistent. In practice you may find that you need to make several minor adjustments to the spreadsheet to produce an easily understandable and meaningful set of results.

The Final Analysis

So, after you have built the spreadsheet, entered all the relevant data and produced a set of results what does it all mean? Well from our example you can see that supplier C has the best BASE CONFIGURATION and most valuable WARRANTY. Yet their system is let down by EXPANSION capabilities, MONITOR quality and the amount of SOFTWARE included so that the overall total is not quite as attractive as supplier A or B. At first sight it appears that our decision should be a straight choice between suppliers A and B with supplier A having a slight edge.

However there is one final item which may significantly influence your decision and that is the subject of cost. You can take the overall points totals and relate them to the actual price quoted by each supplier for their particular configuration (see Figure 2). In this way you can see which supplier is offering the best overall deal in terms of value for money. Always bearing in mind the suitability of a particular system to your requirements.

In my example both systems from suppliers A and B are a pretty good match for our requirements but there is a significant cost differential which makes the system from supplier B the more appealing. This is despite the fact that results for supplier A are a slightly better fit. If the actual cost of a system is not a deciding factor then you can merely omit this cost/value calculation from the method and you can base your final choice on the normal totals calculated so as to achieve the best fit for your requirements.

Conclusion

This method of formally evaluating competing quotations may not be completely foolproof because you face the basic problem of deciding the relative merits of individual components. However, it is a valuable approach which will focus your mind on the areas of the decision making process that really matter. Once you have mastered the technique and used the method a few times you will begin to realise its value in helping you to make objective decisions.

The method can be used to evaluate any number of requirements from any number of suppliers and the more complicated the requirement, for example for a Local Area Network, the more useful this approach is in ensuring that the most critical factors have an influence on the overall decision. Indeed the approach is completely flexible and need not be confined to deciding which PC, server or network to buy. You could use it to decide what car, what Hi-Fi or even what house to buy. Provided you can obtain meaningful quantitative data the method will work.

At the end of the day the decision to buy a particular system may be overwhelmingly influenced by a totally subjective factor such as the style of the system unit. In which case the use of this method may prove to be academic, but at least you will know why you have purchased a particular system and which specific factors have influenced your decision.

Finally if you decide to try out this method, at worst it will focus your attention on the things that really matter rather than on persuasive glossy advertising. At best it will help you arrive at a well considered decision.

PCSA

The Author

Dave Stott is a writer and consultant on systems management. He can be contacted at flingmo@cix.compulink.co.uk.

Understanding PC Buses

Julian Moss concludes his article on the ins and outs of PC bus architectures.

During the CPU/bus master phase, the CPU takes it in turns with a bus master device to access the bus. Where more than one bus master is active, the devices take their turn on a round-robin basis.

During the DMA phase, another round-robin passes control in turn to DMA channels 4, 5, 6 and 7. DMA 4 is the channel used to cascade DMA 0 to 3 through to the bus, so these four channels have in effect a lower priority than the upper three.

Signalling

A simple signalling protocol is used by bus mastering devices. When a master requires control of the bus it asserts MREQx, a slot-specific signal line where x is the slot number. When the card gets its turn to have control of the bus, the ISP informs it of this by raising MAKx.

Future developments like Fast EISA could theoretically provide increased transfer rates of up to 133 MB/sec. However, EISA motherboards and expansion cards are complex - and hence expensive - to manufacture. In view of the competition from faster, cheaper local bus designs, it is questionable whether these EISA enhancements will see the commercial light of day.

VESA Local Bus

Windows 3.0 provided the impetus to develop a faster expansion bus. Existing buses were simply too slow for graphical software using traditional display adapters where the value of every pixel must be transferred across the bus.

One solution to this problem would have been to put the display adapter on the motherboard where its frame buffer memory could be accessed the same way - and as quickly - as the system's main RAM. However this would have removed the advantage of being able to upgrade the adapter at a later date. Instead, manufacturers started to build local bus expansion slots into their motherboards.

Early local bus systems were not satisfactory. Every manufacturer used its own design of slot which meant that buyers were still limited to using display cards that would fit: this usually meant cards made by the motherboard manufacturer. To bring some order to the situation, a group of manufacturers got together under the auspices of the Video Equipment Standards Association (VESA) to agree a standard for a local bus slot. There was a split. One group of manufacturers wanted a solution that could be implemented quickly to meet consumer demand; the others, led by Intel, supported a more elegant solution that would take somewhat longer to develop.

The quick and simple solution became known as the VESA Local Bus or VL-Bus. It is based closely on an early local bus design by chip set maker Opti. VL-Bus is essentially an unbuffered set of data, address and control signals connected directly to the 486 microprocessor. Buffering can be added to increase the load that can be supported, but propagation delays then limit the maximum speed to 33 MHz.

VLB Speeds

VL-Bus normally operates as a synchronous bus running at the same clock speed as the processor: 25 MHz for a 486SX/25 or 486DX2/50, 33 MHz for a 486DX/33, 486DX2/66 or 486DX4/100 and so on. This is an important fact to bear in mind when considering the data throughput. Most comparisons for VL-Bus are based on a speed of 33 MHz, the speed of what is currently the most popular 486 processors and also that of the PCI bus. But the speed may be lower or higher.

The advantage of higher speed is that higher data transfer rates are possible. The disadvantage is that the number of peripherals that can be plugged into the bus is limited. VL-Bus runs up against physical laws: the faster a CPU, the smaller the electrical

> *"The length of time that a PCI master can have control of the bus is determined by both the GNT signal and a latency timer. While GNT is active, the master may keep control of the bus. However, once GNT is removed, the device must relinquish control as soon as the timer has expired."*

load it can support. And expansion slots present a greater load than motherboard-mounted devices.

In the original VL-Bus specification, unbuffered VL-Bus designs were recommended to allow only two devices and two expansion slots at 33 MHz, only one slot at 40 MHz, and none at all at 50 MHz. Buffering allows three devices and three slots at a top speed of 33 MHz.

Another disadvantage of higher bus speeds is that it is more difficult to maintain accurate signal timings. For a given bus operation, signals must appear in a particular order, with sufficient time between them to allow expansion boards to respond. If these timings are not adhered to, a board may not work. Many VL-Bus boards running at higher speeds are not reliable.

Bus Signals

VL-Bus uses a 112-pin MCA-style connector mounted inline with an ISA, EISA or MCA slot. With its limited number of allowable slots, VL-Bus is designed to supplement, rather than replace, existing expansion buses. The design means that boards can take advantage of the features of the other expansion bus, such as software configuration and shared, level-sensing interrupts, where they are supported. In fact, the VL-Bus connector contains only one IRQ contact - IRQ9.

The connector includes 32 data lines (D0 - D31) and 30 address lines (A2 - A31), the same as the 486 CPU. Memory is addressed as an array of 4-byte words. The first two address bits are sent over two of the multifunction pins BE0 - BE3. These pins are also used to control 8- and 16-bit operation by indicating which are the active bytes during read and write cycles. Five more pins ID0 to ID3 are used during system initialisation to convey information about the type of processor, its speed and whether no-wait-state writes are supported.

Cycles

Normal VL-Bus data transfers take two bus cycles each, which equates to a maximum transfer rate of 66 MB/sec at 33 MHz. The bus also supports burst mode using signals BRDY and BLAST. During a burst mode transfer, four 32-bit words can be moved in five bus cycles, giving a transfer rate of 106 MB/sec. However, for larger blocks of data there will be some additional overhead, making average throughput somewhat less than this.

VL-Bus also supports bus mastering, though this was not a feature of the original Opti implementation and early local bus motherboards may not support it. When a device wants control of the bus it asserts the signal LREQ. The bus controller signals its acceptance using LGNT. However, the arbitration logic used is that of the standard expansion bus.

VL-Bus 2.0

The original VL-Bus specification was intended very much to meet the needs of the time. It was not designed for use with non-486 processors such as the 64-bit Pentium. Nor is it ideally suited, in practice, for use with processors running at more than 33 MHz, which now include 486-compatible chips from AMD and Cyrix. These inadequacies are addressed in VL-Bus 2.0.

The new bus specification supports a 64-bit data path. The same connector will be used; the extra data bits being multiplexed on to the address bus. Two new signals are defined which indicate when a 64-bit transfer is in progress. VL-Bus 2.0 should be backwards compatible with existing 32-bit VL-Bus cards.

By reducing capacitance in the bus and tightening up signal timing tolerances, VL-Bus 2.0 permits up to three expansion slots when running at 40 MHz and up to two at 50 MHz.

Caching

Another enhancement is support for write-back caching, a feature of the Pentium processor and also of Cyrix's 486 chips. All bus masters must support a new signal, WBACK#, which is used to inform a VL-Bus master that the data it wishes to access has been modified in cache by the CPU. The device must then wait for the CPU to update main memory with the changed data.

A 50 MHz 64-bit VL-Bus promises some impressive figures - a peak of 400 MB/sec using burst mode. However, real-world throughput may be lower. In practice, it may turn out that VL-Bus 2.0 offers no real performance advantage over the other local bus design, PCI.

PCI

Peripheral Component Interconnect was designed to provide local bus levels of performance, but it also had more ambitious aims, which is why the design took somewhat longer to finalise.

Developed by a committee, of which the main members were Compaq, Digital, IBM, Intel and NCR, it is now an industry standard controlled

> *"Early local bus systems were not satisfactory. Every manufacturer used its own design of slot which meant that buyers were still limited to using display cards that would fit: this usually meant cards made by the motherboard manufacturer."*

PC Buses

> *"PCI looks a better choice. Technically sophisticated and forward-looking, catering for as yet unannounced processors and low power chips, it is a complete bus in its own right. It makes VL-Bus look like a kludge."*

by the PCI Special Interest Group.

The original PCI specification was deficient in a number of ways, including a standard for a connector. This was addressed in revision 2.0 which was released in April 1993.

PCI is not a local bus, though it operates at local bus speeds up to 33 MHz. It is connected to the processor's bus via a bridge. The signals on the PCI bus are not processor-dependent. Members of the PCI SIG include Digital and Apple, and PCI slots will become commonplace in non Intel-based systems made by those vendors and others.

The PCI slot is completely self-contained: PCI boards do not need to access signals from another bus. The physical design places the slot on the opposite side of the mounting bracket to existing PC boards: this allows manufacturers to save board space by making shared slots that can be occupied by either a PCI or a traditional bus board.

PCI boards can be as little as seven inches (17.5cm) long, unlike VL-Bus boards which must be long enough to extend beyond an ISA, EISA or MCA slot. It would be practical to build PCI slots into portables. And PCI looks ahead to the new breed of low-power systems by catering for both 5 volt and 3.3 volt logic.

Loads

Like VL-Bus, PCI is designed to supplement existing expansion buses. The PCI bus can support 10 loads. Motherboard devices are one load each, while slots are two loads. In practice, this means that PCI systems will usually have three PCI slots, leaving four loads free for things like graphics and SCSI controllers.

One of the loads on the PCI bus will normally be a bridge to an ISA, EISA or MCA expansion bus. This will allow the use of existing add-ins for non performance-critical uses. One curious result of this is that you might see ISA expansion slots provided on systems powered by non-Intel CPUs, such as the Digital Alpha.

PCI has been engineered to make board designs simpler. PCI chips are directly connectable to each other without requiring "glue logic", reducing the component count on each board. Once the economies of volume production have been achieved, PCI expansion boards should be cheaper to make than EISA, MCA or VL-Bus boards. If this turns out to be true, there are obvious implications for the future of these other buses.

Because it is not a true local bus, PCI can implement features that would not otherwise be possible. For example, although the 486 processor bus does not support burst writes, the PCI bridge can. The bridge can buffer data written by the 486 and then burst it across the bus; this is estimated to give a 10% improvement in real-world performance compared to VL-Bus at the same clock speed.

The Connector

The 32-bit PCI connector has 120 contacts (the 64-bit slot has 184). Of these, 49 carry digital signals, the rest are supply voltage and ground lines. Data and address information are multiplexed onto the same set of contacts to save space. Every signal is adjacent to a supply or ground line, which helps to keep electromagnetic radiation to a minimum.

PCI is designed to be plug and play. The bus supports three types of address space: memory, I/O ports and configuration space. Configuration space is a 256-byte area within each PCI device which contains information about the type of device, its I/O registers, ROM location and so on. At start-up, the PCI BIOS scans all the devices in the system, resolves address space conflicts and allocates each device a unique IRQ level.

PCI is also designed to be processor independent. One potential obstacle to this is where a PCI board contains ROM code to initialise the device. This has been foreseen. A ROM may contain multiple code images. Information about them is stored in the configuration space. At start-up, the BIOS copies the appropriate code into RAM and executes it. ROM code may be written in native machine code or a hardware-independent metalanguage called OpenBoot - currently a draft IEE standard.

Master-target Architecture

PCI uses a master-target architecture: the device initiating a transfer is the master, the subject of the transfer is the target. All PCI data transfers are performed as bursts to a consecutive series of memory locations beginning at a specified address. Each transfer takes one clock cycle. Thus the peak transfer rate for 32-bit PCI is 133 MB/sec; for 64-bit PCI it is 266 MB/sec. Bursts are not limited in length, though a latency timer on each device - programmable at start-up - prevents a device from hogging the bus.

At the start of a data transfer the Command/Bus Enable signal lines (C/BE0 to C/BE3) specify the type of burst activity. This can be memory read or write, I/O read or write, configuration read or write, plus some special commands. At the same time, the start address is placed on the bus. This is followed by the data phase of the cycle, in which data is transferred in sync with the bus clock CLK. During this phase the C/BEx signals indicate which bytes of the data bus contain valid data, since a transfer might not use the full bus bandwidth.

Bus Mastering

PCI supports bus mastering. Arbitration is a function of the bus control logic. A PCI master requests access to the bus by asserting the REQ signal, and receives permission to take control in the form of a GNT signal. The arbitration process can take place while another data transfer is going on, eliminating some wasted bus cycles.

The length of time that a PCI master can have control of the bus is determined by both the GNT signal and a latency timer. While GNT is active, the master may keep control of the bus. However, once GNT is removed, the device must relinquish control as soon as the timer has expired. The latency timer is programmable, allowing performance-critical devices to receive a larger share of access to the bus.

PCI provides some enhanced features for an ISA expansion bus bridged from the PCI bus. Faster EISA-style DMA is supported, improving the performance of ordinary ISA cards. There is also a new even faster mode DMA Type F. This will give transfer rates of around 8 MB/sec, and is already being exploited by hard disks that use the Enhanced IDE interface.

Conclusions

With five types of expansion bus to choose from, the question is: which one to go for? Board manufacturers must be considering this, too, as it cannot be economic to produce five different versions of every board.

MCA and EISA are firmly established. Offering a higher throughput than ISA, plus greater reliability and properly arbitrated bus mastering, they have been the obvious choice for critical applications like file servers. But both bus types are expensive to manufacture, and offer lower performance than cheaper local bus solutions. Support for MCA and EISA is not going to disappear overnight, but their long-term viability seems doubtful.

ISA, despite its shortcomings, is still the most popular expansion bus. Local bus provides a way round one of its limitations - low throughput - while the Intel/Microsoft Plug and Play initiative, which brings automatic software-controlled setup to ISA boards, solves another: that of difficult configuration. ISA won't go away.

Perhaps the hardest decision is: which local bus to standardise on? VL-Bus has a year's head start and consequently is supported by the greatest range of products. But PCI is catching up fast.

VL-Bus adds a fast 32-bit channel to an existing bus, but it eliminates none of that bus's other shortcomings. A VL-Bus/ISA card still needs manual configuration and uses edge-triggered interrupts, and the motherboard has no proper arbitration for bus masters. The original aim of VL-Bus was to provide a cheap way to boost graphics performance for low-cost PC clones. It is not the best technology for the corporate file server.

PCI looks a better choice. Technically sophisticated and forward-looking, catering for as yet unannounced processors and low power chips, it is a complete bus in its own right. It makes VL-Bus look like a kludge.

PCI is a new technology. Early PCI products have experienced compatibility problems. But only PCI has the potential to become a robust, high performance expansion bus for tomorrow's demanding applications. Coupled with an ISA bus, PCI offers the performance and expansion potential PC users are likely to need, now and in the future.

> *"PCI is a new technology. Early PCI products have experienced compatibility problems. But only PCI has the potential to become a robust, high performance expansion bus for tomorrow's demanding applications."*

PCSA

The Author

Julian Moss is a PC consultant and writer. He can be contacted via pcsa@oakworth.demon.co.uk

From OS/2 To Warp

What is OS/2, and is it better than Windows for your workstations and servers? Is it worth upgrading to the new Warp version? Terence Green has the answers

OS/2 is a protected mode operating system for Intel 386 or better PCs. It exploits Intel's 32-bit architecture to pre-emptively multitask DOS, Windows, and OS/2 applications. Some versions can manage this in as little as 4 MB RAM.

OS/2 is the only alternative to Windows and DOS that you can buy and use today on the average desktop PC. While OS/2 exploits Intel's 386 architecture to a greater extent than Microsoft Windows, IBM has made a poor job of marketing it for some years. Since early 1994, following a realignment instigated by IBM Chairman Lou Gerstner the status of OS/2 within IBM has been more clearly defined and a coherent strategy for OS/2 has been allowed to emerge.

In a nutshell, Gerstner has bet the future of IBM on OS/2 (or rather on the WorkPlace OS (WPOS) technology which flowed from OS/2 development) and PowerPC processors. Over the medium term IBM systems from the AS/400 down will migrate to WPOS technology and the PowerPC microprocessor range being developed in conjunction with Apple and Motorola.

Work which started in the late 1980s aimed at creating a platform-independent "portable" OS/2 resulted in WPOS technology. Based on a MACH microkernel, WPOS will bridge the transition to Taligent, the object-oriented operating system currently under development by Apple, IBM and Hewlett-Packard (see box).

PowerPC

Announced in 1991, the PowerPC is derived from the POWER architecture which IBM developed for its RISC System/6000 UNIX workstations. The 1990 RS/6000 systems represented IBM's first successful commercial implementation of Reduced Instruction Set Computing (RISC) which John Cocke of IBM invented in 1970.

IBM and Motorola will ship a range of Power PC microprocessors, chip- and board-level products, and complete systems through OEM and retail channels.

The first IBM-developed WPOS for the PowerPC will be based on OS/2 2.x. OS/2 for the PowerPC, which should ship in 1995, will run DOS and Windows applications and 32-bit OS/2 applications. 32-bit Intel OS/2 applications will run on PowerPC OS/2 systems when re-compiled.

OS/2 Today

Until just a few weeks ago OS/2 was available in two packages, namely OS/2 2.1 and OS/2 for Windows.

OS/2 2.x is the IBM-developed base product now at version 2.11. It includes complete DOS and Windows emulations. You do not need any DOS version or Microsoft Windows software installed on your PC in order to run Windows and DOS applications. If the ability to run DOS and Windows applications isn't a requirement OS/2 2.11 can be installed without the DOS and Windows support.

OS/2 2.11 can also be installed alongside an existing DOS and Windows installation. IBM offers a couple of schemes which can be used to boot into either DOS/Windows or OS/2 2.11. OS/2 2.11 also offers the option of booting a disk image of DOS within an OS/2 session window. This is occasionally used to support version-specific software or hardware drivers.

OS/2 for Windows includes the DOS emulation but not the Microsoft Windows emulation. This is done to reduce the cost of the product by removing the need to pay Microsoft a royalty for use of Windows source code in the Windows emulation.

If you wish to run Windows applications under OS/2 for Windows you must install, or have installed, Microsoft Windows 3.1. OS/2 for Windows loads the native Microsoft Windows software to run Windows applications. You can use OS/2 Dual Boot or Boot Manager to boot into either DOS or OS/2 to run Windows 3.1 if necessary - any changes made to the Windows 3.1 setup under either system will persist.

OS/2 for Windows can be installed and used with or without Windows support. You can install OS/2 for Windows without having Windows 3.1 available and then later install Windows 3.1 and add support for it to OS/2 for Windows.

Currently OS/2 for Windows differs from OS/2.11 in only one respect - it doesn't include a Windows emulation. Apart from that the package contents are identical. All of the features mentioned below apply equally to both products.

Future enhancements to OS/2 will include the provision of security hooks for third-party add-ons, support for Silicon Graphics OpenGL 2D and 3D graphics, and peer network support. An OS/2 for SMP (Symmetrical Multiprocessor) systems option is available now.

In October 1994, IBM launched a new version of OS/2 which is officially known as OS/2 Warp version 3.0. This product comes in four varieties, not all of which are available right now.

The remainder of this article looks at OS/2 version 2.x, both in its Windows and Window-less versions, as this is the version that is currently in use within most companies and the one that most companies will continue to buy until the more powerful versions of Warp are released. For details of the Warp range, see the History of OS/2 box.

OS/2 In Use

OS/2 (both 2.11 and "for Windows") runs on IBM and IBM-compatible PCs and supports a wide range of peripheral hardware. It has been tested on over 1,000 PCs from many manufacturers. On the peripheral side SCSI adaptor support is comprehensive but SVGA video display support can be tricky with older video cards and, conversely, very new video cards for which Windows drivers normally appear first. While most video adaptors are supported at 1024x768 with 256 colours, only a few currently support 64K and 16.7 million display colours under OS/2.

CD-ROM

OS/2 includes an installable CD-ROM file system and support for most common SCSI and non-SCSI CD-ROM drives. CD-ROM/XA and PhotoCD are supported where available. Multimedia facilities including audio capture/playback, CD-ROM audio and software-based digital video are bundled in the base package. Supported sound cards include IBM products and Creative Labs SoundBlaster and MediaVision Pro Audio Spectrum 16 adaptors.

Advanced Power Management (suspend/resume on notebooks) and PCMCIA card support is also provided by OS/2. Popular notebooks and PCMCIA cards are included in the base support. Others may require additional manufacturers' drivers.

Adobe Type Manager for Windows and OS/2 is part of OS/2. REXX, a powerful command language with programming capabilities is included with OS/2. Support and upgrades are supplied via CompuServe, IBM bulletin boards, and the Internet.

OS/2 has has an object-oriented shell called the WorkPlace Shell which provides a desktop on which files, programs, and network and printer objects can be organised and activated with drag and drop mouse actions.

Disk Consumption

A full installation takes about 30 MB of hard disk space to which must be added a swapfile which averages about 10 to 20 MB for a PC with 8 MB RAM running a handful of applications. The actual size depends on the amount of physical RAM in the PC and the memory required by the application working set.

OS/2 is also the base operating system for numerous IBM networking and connectivity products which must be purchased separately. Full client/server network connectivity is available in the form of LAN Server, Communications Manager (IBM host

History Of OS/2

User acceptance of OS/2 2.x has always suffered from the many mistakes that blighted OS/2 1.x from the launch in April 1987 when IBM tried to pair OS/2 and PS/2, the proprietary hardware technology with which IBM tried and failed to wrest back control of the IBM PC clone market.

Actually shipping OS/2 1.0 eight months late in December 1987 and with no graphical interface wasn't a promising start. It wasn't until OS/2 1.3 shipped in October 1990 that OS/2 started to look interesting as a serious platform but by then it was only as a base operating system for networks, not a desktop contender.

OS/2 has been the sole responsibility of IBM since 1991 when Microsoft dropped OS/2 in favour of Windows and Windows New Technology (NT). OS/2 2.0 had already been demonstrated at Microsoft/IBM developer meetings as early as 1990.

Instead of finalising OS/2 2.0 then, IBM dropped much of the code, rewrote huge chunks in C, and then in an about-turn decided in mid-1991 that when OS/2 2.0 shipped it would support Windows applications. The last was a belated recognition of the massive success of Windows 3.0 but it further delayed OS/2 2.0.

In May, 1992, five years after the announcement of OS/2 promised a new future for DOS users a 32-bit version of OS/2 that could conceivably fit the bill had arrived.

OS/2 2.0 which shipped in May 1992 was actually a hybrid of 16-bit and 32-bit pieces and it had lots of problems, both with hardware and in its Windows emulation. On the plus side it maintained compatibility with applications written for OS/2 1.x.

OS/2 2.1, released in May 1993, had vastly enhanced hardware compatibility with popular PCs and peripherals and fixed most of the Windows problems. However the installation and configuration routines remained as a stumbling block for non-technical users.

OS/2 Warp 3.0 was launched in October 1994. There are 4 versions, all of which are supplied with a Bonus Pak containing some applications such as IBM Works (previously known as Legato), and all of which include internet access as standard via support for PPP and SLIP protocols. OS/2 Warp 3.0 boots in 4 MB. It doesn't include Windows support, so if you want to run Windows applications you'll need a real copy of MS Windows. OS/2 Warp 3.0 Full Pack is a version of OS/2 Warp 3.0 that can run Windows applications without the need for a copy of MS Windows. These 2 products compete with Windows NT Workstation.

OS/2 Warp LAN Client, which was announced in October 1994 and is due to ship in the next month or so, consists of OS/2 Warp 3.0 plus LAN requesters, agents, and other management utilities. It is available both with and without Windows emulation. This product will not boot on a 4 MB machine - it's likely to need around 8 MB.

It is Warp LAN Client that most companies will want to go for, as it includes the necessary components for connecting to corporate networks. Until Warp LAN Client is generally available, IBM (and PCNA) recommends that you continue buying 2.1.

Note that the current version of OS/2 for Windows only supports Windows 3.10. Warp will support 3.11 and Windows for Workgroups.

OS/2 To Warp

connectivity), Database Manager (client/server relational database) and TCP/IP software. Multimedia audio and video software is packaged in the Ultmedia product range, Symmetric multiprocessor systems are catered for by OS/2 for Symmetrical Multiprocessing.

Performance

OS/2 2.11 will install and boot on a PC with 4 MB RAM but it needs 4.5 MB of memory for the operating system alone so it is guaranteed to thrash the hard disk even before any applications are started (OS/2 Warp 3.0 will fix this - see later).

OS/2 for Windows is less demanding. A 386SX is enough to run OS/2 but most users will want to start with at least a 486SX with 8 MB RAM for a well-used desktop PC.

Long filenames are supported in DOS, Windows, and OS/2 sessions but there is no translation for applications which are not aware of long filenames or when running booted from DOS.

OS/2 provides a High Performance File System (HPFS). HPFS is faster and imposes less overhead in terms of cluster allocation than FAT but requires more RAM and works best on large disks, say over 100 MB.

There are two disk compression products for OS/2, both available from third parties. DCF/2 supports FAT and HPFS file systems while Stacker for OS/2 only supports FAT at this time although HPFS support is always said to be under consideration. Stac and IBM recently signed an agreement to replace the compression utility in IBM's PC-DOS with the Stacker compression and it is likely that a future version of OS/2 will include Stacker for DOS and OS/2 with HPFS support.

Performance

OS/2 performs well when compared with DOS or Windows and it is definitely more robust than either. OS/2 2.11 offers very good support for DOS applications including tricky DOS games. For Windows applications the Windows emulation (called WIN-OS/2) in OS/2 2.11 is slightly

Taligent - The Future Of OS/2

IBM and Apple announced their joint development of Taligent in 1991 at the same time as their joint development of the PowerPC with Motorola. Taligent's starting point was a future Macintosh operating system codenamed "Pink" on which Apple had been working alone.

IBM and Apple formed a new jointly-owned company charged with producing a next generation operating system. A little while later Hewlett-Packard picked up 15% of the shares in Taligent and joined the development effort.

The Taligent operating system consists of a microkernel and object-oriented applications and services frameworks. Services frameworks support all shared services such as file systems and networking. Applications frameworks provide the remaining services such as the user interface and communications with shared services.

Taligent object-oriented frameworks support the concept of re-usable code parts, reducing the amount of original programming required and therefore also the cost and time taken to develop and maintain applications.

The frameworks run on top of the Taligent microkernel which by design will isolate the frameworks from the underlying hardware. The microkernel is the only part of the operating system that can pre-empt any other task. It comprises the minimum services required - inter-process communication, memory and I/O management, and task management.

The microkernel provides these services but does not make any decisions about how they are used - that policy is left to the operating system - in this way the microkernel can equally support a powerful workstation operating system or a small handheld computer. Taligent and WorkPlace OS are based on the same microkernel, the Mach 3.0 microkernel developed by Carnegie-Mellon University. (NextStep is based on the Mach 2.5 microkernel.)

In theory a WPOS such as OS/2 for the PowerPC will appear all but identical to Taligent in due course but the Taligent OS will have a pure object-oriented file system while WPOS still plans to provide backwards compatibility with 16-bit DOS and Windows applications. Taligent frameworks have already been demonstrated running on an early version of the Taligent OS and on OS/2 and AIX.

The strategy IBM is outlining is one that allows developers to begin development now on applications with the promise of a seamless transition as WPOS and Taligent arrive. Programmers can begin coding programs for OS/2 now that will run on Taligent. What's more they can develop such applications using object-oriented tools which are language-neutral and they can produce re-usable code parts which through DSOM can be used in distributed applications.

Any application developed for OS/2 for the PowerPC (a WorkPlace OS) should run on Taligent too provided they are written to the full 32-bit OS/2 API or ported from 16-bit OS/2. A recompile should only be necessary if moving between hardware platforms. OS/2 applications developers can write applications now for OS/2 2.x (Intel-based) and expect them to run on the PowerPC and eventually, perhaps in 1996, on Taligent.

Development and porting tools are available now for 32-bit OS/2 applications. In July 1994 IBM made available an alpha version of the OpenDOC developer's kit for OS/2. OpenDOC is a specification for compound documents. It is an open architecture for application development in a distributed environment originated by Apple but now supported by IBM, Novell and others. OpenDOC is not tied to a particular platform being language-neutral and hardware-independent.

faster (by about 10% on average) than the native Windows 3.1 that OS/2 for Windows runs.

DOS memory management nightmares are obsoleted and troublesome applications can be run in protected memory space within virtual DOS machines (VDMs). All current versions of Microsoft Windows run Windows applications in a single VDM hence OS/2's claim that it offers crash protection by making it possible to run individual Windows applications in separate VDMs.

The downside of separate VDMs is that they require more physical memory and swap space and OS/2 only supports OLE 2.0 function between Windows applications running within a single VDM - OLE 2.0 applications running in separate VDMs cannot interoperate.

Another problem that may affect OS/2 users is the single input queue. All applications receive input through a single queue and therefore can potentially stall OS/2 necessitating a system reset. In practice this rarely causes problems on established systems.

Backwards Compatibility

Because there are few horizontal OS/2 applications compared with the number of vertical market OS/2 applications users often run a mix of DOS, Windows, and OS/2 applications and it's here that OS/2 excels because of its compatibility with DOS and Windows device drivers which are simply installed into DOS and Windows sessions under OS/2 just as they would be in native DOS or Windows.

Windows NT, because of its security model, is not able to make use of existing DOS and Windows device drivers and its backwards compatibility with DOS and Windows applications is therefore less comprehensive than that of OS/2 2.11.

Although OS/2 2.11 supports Windows 3.1 in both Standard and Enhanced mode, it shares with Windows NT an incompatibility with Windows VxDs (virtual device drivers). For this reason OS/2 2.11 cannot run Windows applications which access the hardware through VxDs such as backup utilities and Windows for Workgroups network facilities.

OS/2 Warp 3.0 adds support for Windows for Workgroups 3.11 (but not the VxD-based peer network function) and Windows 3.11. The Warp product (OS/2 names are always drawn from Star Trek - the code name for 2.11 was Klingon) also introduces support for Win32s, the Microsoft API developed to support 32-bit applications which can run on Windows 3.1 and Windows NT.

OS/2 For Windows

With Win32s support now in OS/2 any Windows application written to the Win32s API will run on Windows, Windows NT, Windows 95 (when it ships), and OS/2. However, very few applications actually use the Win32s Windows libraries. Win32s has largely been superseded by Microsoft's push to get developers writing to the Win32 API instead so that their applications run on Windows 95 and NT. Although Win32s hasn't been discontinued it isn't promoted by Microsoft any more. Microsoft has said that it will discontinue all 16-bit application development when Windows 95 ships.

The Next Generation

OS/2 Warp 3.0 (see the History of OS/2 box for a rundown) is based on OS/2 for Windows. It features a huge improvement in performance for users with low specification PCs containing 4 MB or 8 MB RAM and a host of usability improvements. Many areas of the WorkPlace Shell desktop have been enhanced following user feedback requesting that it be made easier to use.

One of the areas which has received most attention is the install process, traditionally an OS/2 turn-off. Install now attempts to detect and configure peripheral hardware. The number of questions asked of the user has been reduced and there is a choice between advanced and easy installation paths. The Easy path involves only a handful of dialog boxes. There is also a new and very comprehensive tutorial.

Warp also includes two components which will enable developers to complete 32-bit OS/2 applications which will run on both Intel and PowerPC OS/2 systems and to develop distributed OS/2 applications. The Presentation Manager windowing system which had been the last component to include 16-bit code is now fully 32-bit. System Object Model (SOM) 2.0 and Distributed SOM (DSOM) provide OS/2 systems with client support for platform-independent distributed applications. DSOM effectively can create a software bus on a heterogenous network.

SOM is a language-neutral application development environment which has been chosen as the transport layer for the OpenDoc compound document architecture principally being developed by Apple, IBM and Novell. OpenDoc is roughly analogous to Microsoft OLE 2.0 but differs in that it is designed to be platform independent and to work in heterogenous networks.

OpenDoc, currently still at the alpha test stage, will be made available for Macintosh, Windows, OS/2 and AIX initially.

In addition to feature enhancements and performance improvements IBM has decided to ship a Bonus Pak with Warp. The Bonus Pak includes an entry-level OS/2 application suite comprising word processor, spreadsheet, database and report writer. The BonusPak also includes CompuServe and Internet access packages, a System Information tool based on NetFinity for OS/2, Faxworks for OS/2, and the Person to Person for OS/2 workgroup application that connects users via a shared whiteboard over a LAN or modem link.

PCSA

The Author

Terence Green is a writer and journalist with a special interest in operating systems.

Understanding Printer Technology

Chris Long presents an overview of some of the most common printing technologies found on today's desktops.

Of all the peripherals we buy and support it is the printer that is the one we rely on most. It's also the one that seems to get the least attention. Of course, a few years ago it was so simple - if you needed to print business letters in anything approaching letter quality, you bought a daisywheel printer. If you wanted to print graphics, tough. Although if you could sacrifice text quality for speed, you bought a dot matrix printer. A bit later, after the arrival of the laser printer, it still seemed simple; everybody thought the laser printer's speed and clarity would send all other printers to the big car boot sale in the sky.

Of course, this being the computer industry, things haven't worked out that way. Yes, laser printers are more popular (and faster and sharper), and yes the daisywheels went away, but dot matrix printers have stubbornly refused to die. The market's not as big as it was, but they're still selling. What's more, both dot matrix and laser printers are being usurped by a new pack of high quality, low priced inkjet printers. While at the same time a raft of affordable (but only just) colour printers has arrived, such as the new Colour LaserJet recently announced (but not shipping worldwide just yet) by HP.

Heritage

The list of printers past and present doesn't stop at laser, inkjet and dot matrix. The heritage of the computer printer goes back much further than the PC. There are barrel printers, chain printers, band printers and others. You'll still find these linked to a small number of (mostly large) computers if you look hard. All of those technologies have given a bit to the printers we use now. Some more than others.

The barrel printer has a cylinder (or print drum) that has lines of characters going across it - usually about 136. Each line contains 136 (in this case) repetitions of the same character.

As you go around the drum, each line is a different character, so the line of upper case "A" is followed by upper case "B" and so on.

In front of the print drum is a line of hammers, each hammer lined up opposite a character. There are 136 hammers, one for each character position. To print, the hammer strikes the paper (and ink ribbon) against the letter behind it.

The software looks at the text it needs to print and works out where each letter is in the line. As the drum revolves the system knows where, in relation to the hammers, each letter or symbol is; so when the "A"s arrive under the hammers it fires off all the "A"s needed in the line. Then come the "B"s and so on. Thus the line of text is built up by printing the letters one at a time.

The barrel printer is similar, in some ways, to the band, train and chain printers, where the characters were on a circular band and spun horizontally around not unlike the teeth of a chain saw. Unlike the band printer, which was literally a band with the characters etched on it, the train and chain printers had separate "print slugs" that made up the horizontal print unit and weren't joined up, making it easy to change worn or redundant characters.

Again the heads were opposite the spinning band and again the software knew where each character was and would be able to hit a specific character as it went by.

Timings

It doesn't take much to work out that these types of printers are quite difficult to set up. Hammer flight times are very important when you consider that some of these chains and barrels spin at 3,600 rpm. The slower the hammer, the later it hits the paper, and a hammer that's too slow results in a character that's placed too high (in the case of a barrel printer) or to the side (for a chain printer).

So it is with this heritage that the current printers have taken over the mantle of leaders. First up is the oldest, cheapest, and noisiest models on the market: the dot matrix.

Dot Matrix Printers

The one major advantage of the humble dot matrix printer is that it is the only choice for printing multi-part forms. Nowadays they even come with adjustable impact settings so you can switch from a lightweight tap for single copies to a more violent punch for four or five.

Dot matrix printers are classified by the number of pins or wires in their print heads. The fastest of them are 9-pin printers, where the pins are arranged vertically in a single column. A 9-pin printer produces fast, slightly blurred prints which are good for draft copies.

Clearly, the vertical resolution of the printer is governed by the spacing of the pins on the print head. To increase the resolution, a 24-pin printer has three columns of 8 pins. Each column is very slightly higher or lower than the previous one. So, by carefully choosing which pin to fire, a greater resolution can be achieved. Some 24 -pin printers give you the choice of a regular fabric ribbon or a one-pass film ribbon for crisper print.

Print Action

Most dot matrix printers assume that all data coming from the PC consists of characters that the user wants to print. So send an ASCII 65 to a dot matrix printer and an upper case "A" will probably be printed.

To send a command to the printer, the codes for the command are usually preceded by a special character (often the Escape character, which has an ASCII code of 27). Escape codes, as the commands are called, allow printer drivers to set fonts, spacing, colours and so on.

The print head, containing the 9- or 24-pins, starts its run from one side of the platen (the print roller), solenoids attached to the back of the pins fire at signals from the processor and hit the ink ribbon against the paper. As the head passes over the paper the processor tells it to fire different combinations of pins to make up the different characters needed.

Compatibility

Compatibility isn't generally an issue for dot matrix printers. Most models can emulate one or more Epson standards, and very few software applications lack an Epson-standard driver. Other popular lowest-common-denominator emulations to look for in hardware and software pairs include the IBM Graphics Printer and IBM Proprinter.

A few newer printers, notably Epson's, offer built in scalable fonts as well as the usual 10 and 12 point elite and pica (comparable to Courier) and serif and sans serif proportional (comparable to Times and Helvetica) fonts, although you'll probably find it more important to make sure there's a Windows 3.1 driver that supports that software's scalable TrueType fonts.

A major advantage of the dot matrix is their cheap running costs: ribbons cost a fraction of the price of laser toner cartridges or inkjet cartridges. Although their output isn't the best, they're OK for drafts, invoices, reports and other documents that won't generally need to impress customers. But they're just not sharp enough for modern business correspondence. Generally speaking they are just the job for people who want reliable running, flexible forms-handling, and low running costs.

Inkjets

The inkjet has been enjoying a phenomenal success over the past couple of years, thanks mainly to HP's DeskJet series. It's clear that the inkjet printer is certainly challenging the laser printer as the favourite desk top printer for light users. The running cost of an inkjet printer is higher than that of a laser, and you'll need special paper for top-quality results, but inkjets are cheaper to buy (much cheaper)

Plotters

In the past, plotters could be defined as machines in which pencils or pens moved across large sheets of paper to create images generated in CAD or engineering software. Today the only part of that definition that still applies is the use of large sheets of paper. While pen and pen/pencil plotters are still an important part of the market, a wide range of newer technologies has arrived.

Plotters can be broadly divided into two categories: vector based and raster based. Vector based plotters, which include pen plotters and pen/pencil plotters, not only represent the original plotter technology, they also were the first colour output devices. Because they draw lines from endpoint to endpoint continuously, vector-plotter images are sharper and often more accurate than those created by raster-based plotters.

Vector plotters are versatile because they can print on virtually any media and in virtually any colour and grey scale using a combination of (usually) eight pens or pencils. The main disadvantage of vector plotters is they tend to be slow when drawing complex files.

In another sub division vector plotters can be divided into flatbed and paper movers. The more traditional flatbed is where the media (usually paper) remains stationary and the plotter moves. Paper movers (also called drumfeed, grit wheel or rollerbed) move both the media and the drawing tools.

Flatbed plotters usually are more expensive than paper movers, but they can print on a wider range of media, including tracing paper and heavy cardboard. Also, unlike paper movers, flatbed plotters don't need pinch rollers to hold the media in place (some of them using a vacuum system to hold the paper down), so they can print to the edges of the page.

Raster Based Plotters

All plotters other than pen plotters are raster based, which means the plotter driver has to convert the vector data from the application to the pixel oriented, raster based device. That process normally slows the initial plot, but any subsequent plots are faster. Even considering the translation time, raster plotters usually are faster than vector.

Direct Imaging Plotters

Direct imaging plotters are similar to thermal fax machines. They burn an image into special thermosensitive media and are reasonably cheap to buy. But because they use special media, the cost per plot is relatively high. Another down side is most direct imaging plotters do not print in colour.

Inkjet Plotters

Inkjet plotters are fast, relatively inexpensive and have a moderate cost per plot when standard media is used. However, some manufacturers recommend using special media for optimal output results, which increases the cost per plot.

Electrostatic

Electrostatic plotters, which lay toner on electrostatically charged media, represent the high end of plotters. Their main disadvantage, apart from high price, is that they require a lot of maintenance.

Printer Technology

> *"You can download printer drivers from bulletin boards run by the big software manufacturers. Companies like Microsoft and WordPerfect have their own systems in some countries and a lot of companies have special areas on CompuServe."*

than lasers. In addition, inkjets have been managing colour printing for years, whereas colour laser printers are only just starting to appear on the desktop. Until now, colour laser printing meant a colour photocopier with an add-on parallel interface, which costs around the same as a luxury car.

Operation

Technically, inkjets work like dot matrix printers. They receive data from the PC and process the data to be printed, but instead of little pins they spray ink from tiny nozzles.

Both of the leaders, Hewlett Packard and Canon, use what Canon calls bubble jet technology, which doesn't actually pump the ink through the nozzles. Instead, the ink is heated, causing a bubble to expand and push a drop of ink toward the page.

Thinner than a human hair, there are up to 256 nozzles in the print head and each is an ink-filled chamber. Just as in the dot matrix printer, a character is made up from a collection of dots of ink on a page and the printer's processor is in charge of creating those dots.

At the back of these ink filled chambers is a heating element. When it receives power it heats the ink to over 900 degrees and the ink immediately boils. The boiling ink is at the back of the chamber and, as it expands, it pushes the cooler ink from the front of the chamber out of the nozzle at the front. This hits the paper and makes up the character. The drop in pressure from the blob of ink leaving the chamber "sucks" up more ink from the ink cartridge so the chamber stays full. The character itself is made up from a matrix of about 20x20 dots.

Continuing the similarity, the inkjets' built-in fonts and use of Windows fonts are similar to those of dot matrix printers. If for some reason you can't use the inkjets' own software drivers, Canons emulate matrix printers, working, for instance, with Epson drivers, while HP's are compatible with LaserJets.

Except for the noise of the paper transport, inkjet printers are practically silent. This may be a consideration if you're looking for a printer to be used in an open plan office and you don't like the fumes and sounds that some lasers produce. Most use standard sheets of paper, though some wide carriage and tractor feed models are available.

Downside

The downside is that inkjets are really quite slow, most of them averaging under two pages per minute for black and white graphics

Those Typeface Terms In Full

ATM - Adobe Type Manager from Adobe Systems is a font utility, first developed for the Macintosh and now equally popular with Windows (and built into OS/2). ATM produces WYSIWYG screen fonts, and, for non-PostScript printers, printer fonts as well, for any and all of the nearly endless list of PostScript Type 1 scalable fonts. An enormous font library is ATM's main advantage over its rival, Microsoft's TrueType (or similar products such as Bitstream's Facelift). ATM's main disadvantage is that it has to be loaded after Windows, whereas TrueType is built into Windows.

Bit mapped font - A fixed set of characters stored in a printer's memory at a specific size and a predetermined resolution, or level of quality (eg, 10 point Courier at 300 dots per inch). A bit-mapped font can be printed only at that size and resolution.

Font - In the world of computers, the terms font, typeface and type style are often used interchangeably to mean a complete set of characters that share the same design such as Times, Courier, Palatino. Then that font has the usual attributes, eg, bold, italic, condensed etc applied to it.

Scalable font - A set of characters that can be printed at any size or resolution. The characters are stored in the printer as descriptions of their shape, allowing users to decide size and resolution.

Soft font - A set of characters that exists as software files on the computer instead of being built into the printer or resident in a plug-in cartridge.

PostScript Type 1 fonts - Most PostScript fonts are known as Type 1 fonts, though a few (Type 3 fonts) take advantage of the language to create fancy fill and pattern effects. PostScript generally comes in two flavours, the original Level 1 and Level 2, which adds support for colour plus other enhancements such as data compression.

TrueType - Created for Apple's System 7 Macintosh operating system and shared by Microsoft Windows 3.1, TrueType is a standard used by a rapidly expanding library of scalable screen and printer fonts. As a bonus, TrueType fonts are portable, meaning users can read or print documents that contain fonts not installed on their (the recipients') PCs.

(and even slower in high-resolution modes) and 100 characters per second for text. The latter is a respectable speed compared to a dot matrix printer's near letter quality modes, but it's no match for the matrix printer's rapid draft mode.

Inkjets also suffer, in varying degrees, from smudging, smearing, and seeping. Their liquid ink may bleed into absorbent paper, resulting in slightly blurred or fuzzy output, and some inks don't set well or dry quickly. With the first HP DeskJets, brushing a finger against the page made documents unreadable. Today's models aren't as vulnerable (newer DeskJets, for instance, use a smudge-resistant oil-based ink). But you still don't want to read inkjet pages in the rain. Also ink cartridges tend to be expensive - and don't last as long as dot matrix printer ribbons.

Refills

As cartridges don't last that long, users sometimes have a go at refilling a cartridge with their own ink. It has a lot of logic to it: it's cheaper, more efficient and more environmentally friendly because you don't have to throw the old one away, but care has to be taken. It generally involves a syringe, needle and a bottle of good quality ink (and if you are going to attempt it, it is worth remembering that it is better to top up a cartridge rather than fill an empty one from scratch).

There are also kits designed to help inkjet users to do this in a more controlled way making the job a bit less messy. But, fundamentally, it has to be remembered that this could invalidate your printer's warranty and possibly irreparably harm your printer. Modern inks are a carefully-controlled mix of dyes and lubricants, and refilling a cartridge with an ink for which it was not designed is just asking for trouble.

Laser Printers

Laser printers don't have the same moving parts as dot matrix printers and inkjet printers. Rather than moving a print head back and forth over each line, they produce a whole page at once. This makes their mechanics, and their control language, more complex.

The first stage in the printing process involves applying an electrostatic charge to a selenium-coated drum. Then, a laser draws the image of the page onto the drum. To achieve this, the drum rotates slowly. Between the laser source and the drum is a glass prism, which also rotates slowly. As the prism rotates, it moves the laser beam across the drum, while the rotation of the drum moves the beam in the other direction. The software inside the printer flashes the laser on and off so that laser light hits all the desired spots on the drum.

The effect of the laser on the drum is to discharge the selected areas, thus leaving an image, comprised of "static charge" on the drum.

Then, the drum rotates and goes through a "toner bath" where the parts of the drum that are charged pick up toner and the parts that aren't charged remain clean. Next paper is pressed against the drum, transferring the toner. The remaining electrical charge on the drum is removed by a "corona wire" (yes, that's what it's for) before it comes back round to the laser beam.

This is all done in one cycle. Meanwhile the paper goes through heated pressure rollers that melt the wax particles in the toner to fix it to the paper.

Some vendors use light emitting diodes (LEDs) instead of lasers, but the results are the same, ie quiet, good-quality prints.

Changes

The laser printer market has changed since HP launched the LaserJet a decade ago. It's getting rare, for example, to find a laser printer that has less than 1 MB of on-board memory (remember that the printer has to build up an image of the entire page in RAM, before transferring it to the drum). Printer RAM is also used to store downloadable fonts, though the more advanced printers have their own hard disks to save on download time.

Besides memory, assembling the

Page Description Languages

CaPSL - While Canon engines are used in most Hewlett Packard and Apple laser printers, Canon's own lasers until recently used something called Canon Printing System Language (CaPSL). It was a full-featured language complete with scalable fonts, but not that popular and didn't get much support in the market. Now more recently Canon has take a more aggressive stance in the market with its CanOPE (Canon Open Printing Environment) printers, which promises compatibility with PostScript and PCL as well as CaPSL.

HPGL - CAD and other design types look for compatibility with the Hewlett Packard Graphics Language, a vector graphics standard developed for HP plotters.

PCL - Printer Control Language, the software standard used by HP's laser printers. The LaserJet Series II's PCL 4 is the lowest common denominator for today's lasers; the more modern PCL 5 includes support for scalable fonts and other improvements introduced with the LaserJet III. HP's latest LaserJet 4 printers use an updated version of PCL 5 that allows 600-dpi printing of graphics and text. Don't confuse PCL with PDL, the generic abbreviation for "page description language."

PostScript - Used to be the only way to get scalable fonts on your printer, now given a run for its money by competing technologies, but Adobe Systems' PostScript is still the king of page description languages, particularly for desktop publishing. PostScript's big advantage is it's a device independent language, that describes complete pages, including text, graphics. So the print-to-disk PostScript file that prints on a LaserJet will also print on a professional typesetting machine so long as it has the necessary fonts (or the PostScript file contains the bit-images of them).

Printer Technology

> *"There are barrel printers, chain printers, band printers and others. You'll still find these linked to a small number of (mostly large) computers if you look hard. All of those technologies have given a bit to the printers we use now."*

data that make up a fancy graphics page takes processing power. Most laser printers feature on board processors, much like the graphics co-processors on video boards. And now we are seeing high end models using powerful RISC processors to make them even faster.

Resolution

Increased speed is only part of today's laser story - another is increased resolution. The finest detail a printer can reproduce is measured in dots per inch, although its true resolution is defined by the number of printer pixels per inch that the printer actually holds in memory before transferring that pixel-image to the page.

The standard resolution of 300 dpi is perfectly adequate for general correspondence and charting, but higher resolution printers are appearing all over the place. The 600 dpi lasers are coming down in price and are now in the same price league as their 300 dpi predecessors. In comparison, by the way, professional typesetting machines handle anything up to 2400 dpi. [*PCSA is typeset at 1200 dpi - Ed.*]

Remember that the leap from 300 to 600 dpi is larger than it seems - we are jumping from 90,000 to 360,000 dots per square inch. That's a four-fold increase in the amount of RAM and processing power required to produce the image.

Some laser printers are upgradable from 300 to 600 dpi or more via added memory and other third-party options. Some vendors use resolution-enhancement techniques, which increase the apparent resolution of a printer by producing partial pixels to smooth out jagged edges. Image enhancement requires less memory than increasing the printer's true resolution, and it can achieve similar results. More interesting for desktop publishers is a similar technique used to enhance halftone images, by supporting additional shades of grey to fool the eye with perceived resolutions of 800 dpi or higher.

Unfortunately, because vendors increasingly confuse true resolution with apparent resolution in their specifications, print samples are the only reliable way to judge output quality.

To judge the quality of a laser printer's output, look very closely at the edges of curved letters such as "e" or "o". Print out the word "Printer" on your printer in a Times typeface, at around 42 point, and compare it with the same word in the title of this article.

Performance

Both Microsoft and Fujitsu have released systems that help you eke a little more performance out of you laser. Laserboost is a hardware upgrade that will replace the system board in most LaserJet II and III models, and some Brother and Canon machines. It is distributed by Pacific Semiconductors and claims to increase print throughput.

The Windows Printing System, from Microsoft, is a collection of fonts and a cartridge for a LaserJet II or III. Again it is meant to increase throughput but is also aimed at giving users more functionality under Windows.

Engines

Despite the bewildering array of printers on the market, there are relatively few laser printer engine manufactures. The advantage of this is that you are likely to be able to use printer supplies like toner or drums from other printer systems. This enables you to shop around and perhaps get a better deal.

An even better deal is to be had by buying refilled toner cartridges instead of brand new ones. These save you money, and are a little more ecologically friendly. Hardly surprisingly the manufacturers frown on the idea, not least because it removes their lucrative cash channel of replacement cartridges.

On the other hand, just like refilling inkjet cartridges, they have a point. You can easily ruin the innards of your laser printer (expensive innards at that) with substandard materials. It is almost never beneficial to replace the toner yourself, especially with photocopier toner, and you are likely to find that the warranty on your printer has just died along with your printer. So it has to be attempted with care.

Save The Planet

There is a move to make the laser printer more environmentally friendly. The Kyocera Ecosys LED printer was one of the first "green printers" available. It emits virtually no ozone and has an amorphous silicon drum that essentially lasts as long as the printer; normally drums are replaced every 50,000 pages.

The Ecosys imaging drum is guaranteed for three years or 300,000 pages and is claimed to be able to handle recycled paper, something most printers balk at.

This means that replacement drums don't have to be made - or disposed of. The toner is specially developed (it is ceramic based) and conditions and polishes the drum during printing. Also, while many manufacturers participate in a toner-cartridge recycling program, Kyocera takes the idea one step further and you just add more toner to the Ecosys's reservoir.

Connection

If you aren't careful, connecting printers can be a problem too. In addition to the usual parallel and serial ports, many mid range and high end lasers come with Ethernet or AppleTalk network interfaces or both. This makes them suitable for sharing among a few people in a workgroup and can be cheaper than buying a printer for each person. Some printers come with both parallel and serial interfaces, and the ability to accept print jobs from each. This makes it easy to share the printer between two (but no more than two) computers.

Unfortunately, sharing printers can be more trouble than it's worth unless your printer offers emulation switching. This is the ability to handle incoming data from different computers using different software (usually the Macintosh favourite, PostScript and the PCL code used by LaserJets).

Colour Printing

Of all the news in the printer world the most interesting has to be the recent boom in affordable, high quality colour printers. Once too expensive for anything except prepress work in professional print shops, colour printers are appearing in offices everywhere.

At the bargain basement end is the dot matrix printer fitted with multicolour striped ribbons. Dot matrix colour is fine for occasional use, but it looks mediocre at best. The colour is delivered by moving the multicolour ribbon up and down in front of the print head and colours are made by mixing the different inks. With often (very) unexciting results.

Colour inkjet printers deliver bright hues at low cost, although they're somewhat vulnerable to banding or smudging, particularly when using plain paper instead of special coated stock, or when printing light and dark colours adjacent to each other. Although some top range machines are equipped with RISC processors that address these problems and are compatible with professional colour matching systems such as the Pantone standard, these are not cheap.

Perhaps the best known colour printing scheme is thermal wax transfer, which applies heat to sheets of waxy ink, fusing the ink to the page. The need for special paper, and the fact that each printed page uses three or four sheets of coloured ink, makes for a high cost per page, but thermal wax colours are rich, crisp, and excellent for transparencies as well as standard pages.

Because of their price the two remaining colour technologies remain relatively rare, but are worth watching. First there is solid ink printing, which works rather like an inkjet: it melts solidified ink and squirts it onto the page where it dries. Unlike most of their rivals, solid-ink printers produce very good colour without requiring special paper.

The other, dye sublimation, is the height of colour printing (with a price to match). The printer heats a solid dye on the transfer roll until it sublimates into gaseous form. The gas is absorbed by a specially coated paper or transparency and is applied at varying intensities to produce multiple hues and shades of colour, so, instead of dots, the process produces continuous tones.

Thus the printer's ability to modulate the heat creates millions of colours without dithering (dithering is mixing adjacent dots of a few colours to create the illusion of additional hues) used by most colour printers (like, say, a thermal wax transfer printer, which uses halftone imaging, or dithering, to simulate mixed colours).

The benefits of continuous tone imaging come to the fore with blended colours such as orange or purple. For example, creating an orange dot the dye sublimation printer will print magenta and yellow on top of each other, so the resulting dot is the precise hue intended. A thermal transfer printer, on the other hand, places dots of magenta and yellow very close together in a pattern so that a particular area looks orange.

In other words, a dye-sublimation page looks very much like a colour photograph, the only down side of the technology being that the diffusion process makes for slightly fuzzier line drawings than thermal printing.

Colour laser printers are rare, though HP (as mentioned earlier) is about to start shipping one. It works by having four different coloured toners (yellow, magenta, cyan and black). The paper is passed through the print engine four times. Each time, a different coloured toner is used.

Drivers

Finally, a word on drivers. It is all very well having the best printer in the world, but if your software can't talk to it you will be better off using it as a doorstop. If you include their emulation modes, most printers will be supported by most software, but the fastest and often most sophisticated support comes in "native mode".

To this end it is worth checking with the software manufacturer to see if there is a specific driver for your printer (assuming it isn't on the distribution disks). And it is definitely worth seeing if there is a newer driver available than the one you have already. This is because manufacturers will often build a driver for a printer when developing the program and then refine it later on. Sometimes the newer driver will be a bit faster or work better with your printer (this is also the same for graphics drivers).

You can also download printer drivers from bulletin boards run by the big software manufacturers. Companies like Microsoft and WordPerfect have their own systems in some countries and a lot of companies have special areas on CompuServe with the latest software drivers available to all.

PCSA

The Author

Chris Long is a writer and editor on computing matters. He can be contacted as chrisl@cix.compulink.co.uk.